

# 低信噪比环境下的语音识别方法研究

王 群, 曾庆宁, 谢先明, 郑展恒

(桂林电子科技大学信息与通信学院, 广西桂林 541004)

**摘要:** 单通道语音信号在信噪比较大的环境下经过增强后再识别, 能表现出较高的识别率。但是在低信噪比环境下, 增强后语音信号的识别率急剧下降。针对此种情况, 提出了一种用在识别系统前端的语音增强算法, 该增强算法将采集到的带噪语音信号先使用对数最小均方误差(Logarithmic Minimum Mean Square Error, LogMMSE)提高其信噪比, 然后再利用改进的维纳滤波去除噪声残留并提升语音可懂度, 最后用梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)和隐马尔科夫模型(Hidden Markov Model, HMM)对增强后的语音信号做特征提取并识别。实验分析结果表明, 该方法能有效地抑制背景噪声并减少噪声残留, 显著提升低信噪比环境下语音识别的准确性。

**关键词:** 语音增强; 低信噪比; 改进维纳滤波; 对数最小均方误差算法; 语音识别

中图分类号: TN912.34

文献标识码: A

文章编号: 1000-3630(2017)-01-0050-07

DOI 编码: 10.16300/j.cnki.1000-3630.2017.01.010

## Research on speech recognition in low SNR environment

WANG Qun, ZENG Qing-ning, XIE Xian-ming, ZHENG Zhan-heng

(School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China)

**Abstract:** The accuracy rate of single channel enhanced speech recognition in high SNR environment is acceptable, but not so in low SNR environment. In this case, speech enhancement based on logarithmic minimum mean square error (LogMMSE) algorithm and modified Wiener filter algorithm is presented. Firstly the gathered speech signals' SNR is improved by the LogMMSE algorithm. Then using the improved Wiener filter algorithm removes residual noise and improves the signal quality. Finally the enhanced speech is used for recognition by MFCC and HMM algorithms. Experimental results show that the proposed method can effectively remove the background noise and reduce the residual noise, significantly increase the accuracy of the automatic speech recognition in noisy environment.

**Key words:** speech enhancement; low SNR; modified Wiener filter; LogMMSE algorithm; speech recognition

## 0 引 言

语音识别主要是指能够让机器听懂人所说的话, 即在特定或非特定环境下准确识别出语音的内容, 并根据识别出的内容去执行相应的操作。语音识别在车载导航、视频监控、网络视讯等人机交互领域有着非常广泛的应用。例如视频监控往往存在较多盲区, 利用语音识别技术可有效识别出视频盲区内外的危险语音信号, 对某些突发情况及时做出反应。在识别系统中, 一般是将理想环境下语音训练出的模型应用于真实的含噪环境中。而在实际环境中, 由于背景噪声的影响, 含噪语音的识别率急剧下降, 甚至还会出现无法工作的现象。近些年来,

理想环境下的语音识别技术发展迅速, 单通道语音识别技术在理想环境中已经达到了较高的识别率。而在低信噪比环境下, 如何提高语音识别率成为人们关注的焦点。近年来在语音识别抗噪方面很多人做了大量研究, 例如谱减算法、维纳(Wiener)滤波、最小均方误差(Minimum Mean Square Error, MMSE)估计等<sup>[1]</sup>。这些算法虽然能有效地去除噪声, 但都会不同程度地产生失真或引入音乐噪声, 反而使增强后的语音识别率更低。这在低信噪比环境下更为明显。近年来有人提出最小控制递归平均(Improve Minima Controlled Recursive Averaging, IMCRA)改进噪声估计的对数最小均方误差(Logarithmic Minimum Mean Square Error, LogMMSE)算法<sup>[2]</sup>。该算法具有一定的增强效果, 但计算量大, 且识别效果并不突出。针对此种现象, 本文通过在识别系统前端先采用基于对数最小均方误差算法提高含噪语音信噪比, 再使用改进的(Wiener)滤波去除噪声残留并提升语音可懂度, 从而提高语音识别率。

收稿日期: 2016-07-20; 修回日期: 2016-09-29

基金项目: 国家自然科学基金(61461011)、教育部重点实验室 2016 年主任基金(CRKL160107)资助项目。

作者简介: 王群(1990—), 男, 湖北随州人, 硕士研究生, 研究方向为语音信号增强、语音识别。

通讯作者: 郑展恒, E-mail: glzzh@guet.edu.cn

# 1 语音增强

## 1.1 谱减算法

谱减算法<sup>[3]</sup>的基本原理为假设信号是不相关的加性噪声，通过快速傅里叶变换(Fast Fourier Transformation, FFT)，从含噪语音中减去噪声短时幅度谱，将计算所得纯净语音的短时幅度谱结合含噪语音的相位，再经过快速傅里叶逆变换(Inverse Fast Fourier Transformation, IFFT)，得到需要的纯净语音信号，噪声的短时幅度谱可以在语音的静音段或者间隙进行重估和更新。基本框图如图 1 所示。

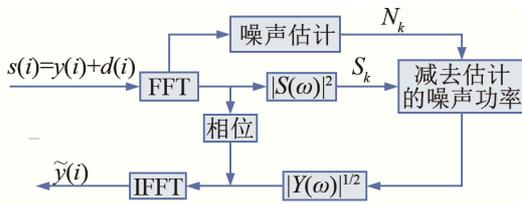


图 1 谱减法框图  
Fig.1 The flow chart of spectral subtraction

图 1 中： $y(i)$  表示纯净语音信号； $d(i)$  表示不相干的加性噪声； $S_k$  表示含噪语音  $s(i)$  的功率谱系数； $S(\omega)$ 、 $Y(\omega)$  分别表示  $s(i)$  和  $y(i)$  的傅里叶变换； $N_k$  表示噪声估计的功率谱系数。相位信息则由含噪信号的相位信息表示。所以经过谱减后的理想噪声信号频谱  $\tilde{Y}(\omega)$  可由式(1)得到：

$$\tilde{Y}(\omega) = [S_k - N_k^2]^{1/2} \quad (1)$$

然后再通过求取  $\tilde{Y}(\omega)$  的 IFFT 可求得最终纯净的时域信号  $\tilde{y}(i)$ 。

## 1.2 对数最小均方误差 (LogMMSE)

最小均方误差<sup>[4]</sup>(MMSE)是基于线性的幅度谱误差平方计算得来。在数学上相对容易处理，但在听觉上却不太适合。于是就有了对数的幅度谱误差平方估计器。logMMSE 噪声抑制算法的主要作用是从已知的含噪功率谱  $\gamma_k$  中估计出纯净语音谱  $\chi_k$ 。由于语音的短时平稳性，可忽略频谱相位对语音产生的干扰，仅对频谱幅度做对数变换。 $k$  表示频率点数。其计算式如式(2)所示：

$$e = E[(\log \hat{\chi}_k - \log \chi_k)^2] \quad (2)$$

通过求  $\log \chi_k$  的条件均值得：

$$\log \hat{\chi}_k = E[\log \chi_k | \gamma(\omega_k)] \quad (3)$$

求解式(3)可以得到估计出的纯净信号幅度值  $\hat{\chi}_k$  为

$$\hat{\chi}_k = \exp(E[\log \chi_k | \gamma(\omega_k)]) \quad (4)$$

对式(4)求  $\log \chi_k$  以  $\gamma(\omega_k)$  为条件的矩量生成函数。再对矩量生成函数求导可得到估计后的频谱幅度：

$$\hat{\chi}_k = \frac{\xi_k}{\xi_k + 1} \exp\left(\frac{1}{2} \int_{v_k}^{+\infty} \frac{e^{-t}}{t} dt\right) \gamma_k \quad (5)$$

其中： $\xi_k$  为先验信噪比， $\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)}$ 。 $u_k$  是后验信噪比， $u_k = \frac{\gamma_k^2}{\lambda_d(k)}$ 。 $\lambda_x(k)$  和  $\lambda_d(k)$  分别表示第  $k$  个频率点的信号方差和噪声方差。求得  $v_k = \frac{\xi_k}{\xi_k + 1} u_k$ 。先验信噪比和后验信噪比影响增益函数<sup>[5]</sup>。

最后，可求得 LogMMSE 估计器的增益函数  $G_{\log}(k)$  为

$$G_{\log}(k) = \frac{\xi_k}{\xi_k + 1} \exp\left(\frac{1}{2} \int_{v_k}^{+\infty} \frac{e^{-t}}{t} dt\right) \quad (6)$$

## 1.3 最小控制递归平均算法(IMCRA)

含噪语音中，噪声的准确估计对语音增强尤为重要。如果噪声估计不准确，会产生音乐噪声或影响增强效果。递归平均型噪声估计通过对带噪语音频谱高频区域所提取的信息进行估计和更新，能避免因噪声更新周期过长而噪声变换过快所导致的估计不准确<sup>[6]</sup>。IMCRA 算法通过引入条件概率  $p(k, l)$  来对噪声进行估计，可通过式(7)得到：

$$\bar{\eta}_d(k, l+1) = \bar{\eta}_d(k, l)p(k, l) + [\alpha_d \bar{\eta}_d(k, l) + (1 - \alpha_d)Y(k, l)^2](1 - p(k, l)) \quad (7)$$

其中， $\bar{\eta}_d(k, l)$  表示第  $k$  个频带的第  $l$  点的噪声功率谱密度估计； $\alpha_d (0 < \alpha_d < 1)$  是一个平滑参数； $Y(k, l)$  定义为第  $k$  个频带的第  $l$  点的信号幅度。式(7)中的条件概率  $p(k, l)$  通过两次平滑和最小值搜索计算得到，计算过程如下：

第一次平滑由式(8)得到：

$$S_f(k, l) = \sum_{i=k-m}^m b(i) |Y(k, l)|^2 \quad (8)$$

其中： $m$  为连续帧个数； $b(i)$  为加权因子。

第二次平滑如式(9)所示：

$$S(k, l) = \alpha_s S(k-1, l) + (1 - \alpha_s) S_f(k, l) \quad (9)$$

其中， $\alpha_s$  为平滑因子，再经过最小值搜索得到式(10)：

$$S_{\min}(k, l) = \min\{S(k, l) | k-m+1 \leq k \leq k\} \quad (10)$$

最后得到条件概率估计值：

$$p(k, l) = \frac{1}{1 + \frac{q(k, l)}{1 - q(k, l)} (1 + \xi(k, l)) \exp(-v(k, l))} \quad (11)$$

其中:  $v(k,l)=\frac{u(k,l)\xi(k,l)}{1+\xi(k,l)}$ ,  $u(k,l)$ 、 $\xi(k,l)$  分别为后验信噪比和先验信噪比;  $q(k,l)$  为先验语音不存在概率, 先验语音语音不存在概率通过平滑和最小值跟踪的两次迭代估计得出。

#### 1.4 改进的 Wiener 滤波

Scalart 在频域 Wiener 滤波的基础上提出基于信噪比(Signal Noise Ratio, SNR)的 Wiener 滤波, 其计算量小且便于实时处理。这种方法需要用到语音的先验 SNR 和后验 SNR。后验 SNR 计算的难度不大, 假设噪声是在平稳信号条件下, 很容易估计出噪声的方差。但是先验 SNR 却不同, 纯净语音的功率在实验中是无法直接得到的。若  $\xi(\kappa, \lambda)$  和  $u(\kappa, \lambda)$  定义为先验信噪比和后验信噪比。再通过判决引导法来估计先验 SNR<sup>[7]</sup>:

$$\xi(\kappa, \lambda) = aG_w(\kappa, \lambda-1) + (1-a) \max[u(\kappa, \lambda)-1, 0] \quad (12)$$

其中,  $G_w(\kappa, \lambda-1)$  表示为前一帧的增益函数,  $u(\kappa, \lambda-1)$  为前一帧的后验 SNR,  $a$  为平滑系数, 一般取小于且接近于 1 的值。

因此, 可以得到在第  $\lambda$  帧的 Wiener 滤波增益函数用先验 SNR 表示为

$$G_w(\kappa, \lambda) = \frac{\xi(\kappa, \lambda)}{\xi(\kappa, \lambda)+1} \quad (13)$$

在 Wiener 滤波中用直接判决法估计先验信噪比会出现高估和低估的情况<sup>[8]</sup>。研究表明, 在 -10 dB 以下的区域存在较多的高估, 在放大畸变大于 6.02 dB 的区域存在较多的低估。高估和低估会导致语音信号增强效果不明显或失真。所以, 可以从两方面对 Wiener 滤波器进行改进。

首先分两步来估计先验信噪比, 第一步估计为式(13), 在(13)基础上进行第二部估计:

$$\hat{\xi}(\kappa, \lambda) = \frac{\xi(\kappa, \lambda)}{\xi(\kappa, \lambda)+1} \left( 1 + \frac{\xi(\kappa, \lambda)}{\xi(\kappa, \lambda)+1} * u(\kappa, \lambda) \right) \quad (14)$$

对于在 -10 dB 以下的区域, 人工引入偏差值修正增益函数, 修正后可表示为

$$\hat{G}_w(\kappa, \lambda) = (G_w(\kappa, \lambda) + c) \frac{1}{1+c} \quad (15)$$

其中,  $c=b/(1-b)$ ,  $b$  为先验信噪比在 -10 dB 以下区域的修正系数,  $0 < b < 1$ , 经实验表明, 当  $b=0.02$  时效果最优。

文献[9]中指出放大畸变大于 6.02 dB 时, 有:

$$E[\hat{G}_w^2(\kappa, \lambda)\gamma^2(\kappa, \lambda)] > 4E[|\chi(\kappa, \lambda)|^2] \quad (16)$$

所以有:

$$\hat{G}_w^2(\kappa, \lambda) \left( 1 + \frac{1}{\xi(\kappa, \lambda)} \right) = 4 \quad (17)$$

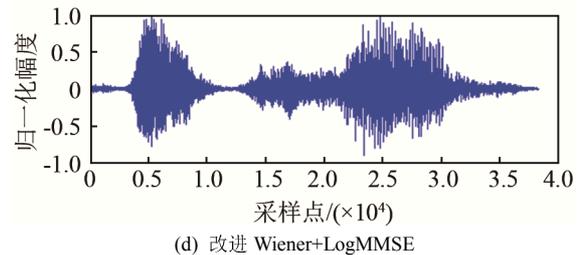
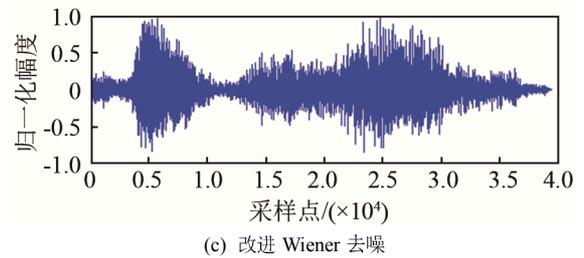
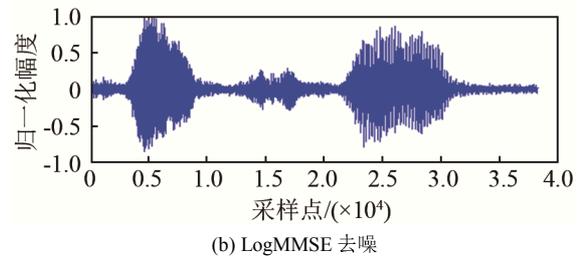
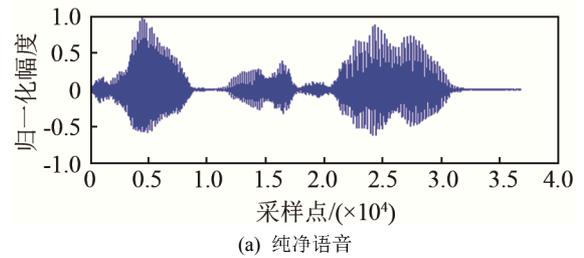
对增强后的语音幅度谱放大畸变大于 6.02 dB 的语音进行限制:

$$\tilde{\chi}(\kappa, \lambda) = d\chi(\kappa, \lambda) \quad (18)$$

其中参数  $d=0.08$ ,  $0 < d < 1$ 。

本文采用对数最小均方误差(LogMMSE)和改进 Wiener 滤波串联形式对带噪语音进行增强处理。先经过 LogMMSE 提高带噪语音信噪比, 再用改进 Wiener 滤波去除噪声残留, 同时还对容易产生畸变失真的区域进行增益补偿, 最大限度地减小因去噪所造成的信号失真, 以此提升识别率。

图 2 为 LogMMSE 和改进 Wiener 在 0 dB、白噪声环境下级联方式的对比分析。其中图 2(a)为纯净语音信号; 图 2(b)为 LogMMSE 去噪效果; 图 2(c)为改进 Wiener 去噪效果; 图 2(d)为先经过改进 Wiener 后使用 LogMMSE 去噪效果; 图 2(e)为本文所使用方式的去噪效果。从图 2 中可以看出, 采用 LogMMSE+改进 Wiener 的级联方式去噪效果最佳。



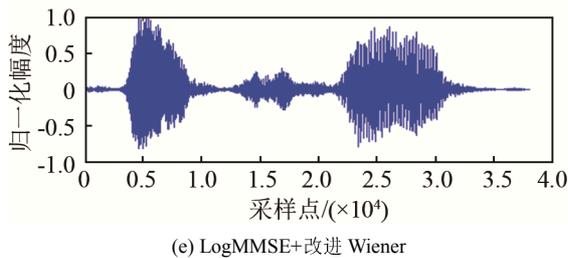


图 2 不同级联方式的对比分析  
Fig.2 Comparative analyses of different cascade modes

从图 2(b)、2(c)图中可以看出 LogMMSE 相比于改进 Wiener 在低信噪比有更好的去噪效果,产生毛刺较少。这样先经过 LogMMSE 处理后再使用改进 Wiener 去除噪声残留效果更为明显。若采用前置改进 Wiener 先对信号做去噪处理,这样得到的信号噪声残留较大,即使最后再使用 LogMMSE 去噪并不能达到理想去噪效果。所以本文采用 LogMMSE 后置改进 Wiener 滤波的级联方式。

## 2 语音识别

本文使用的识别系统为基于隐马尔科夫模型 (Hidden Markov Model, HMM) 的非特定人语音识别。识别系统提取增强后语音的声学特征,再通过维特比 (Viterbi) 算法解码匹配到最佳状态序列得到识别结果。该识别系统采用梅尔频率倒谱系数 (Mel-Frequency Cepstral Coefficients, MFCC) 做特征提取,采用隐马尔科夫 (HMM) 得到训练模型。

### 2.1 梅尔倒谱系数 (MFCC)

MFCC 由于更符合人耳的听觉特性,因此在语音识别中得到了广泛应用。经过采样后的语音信号,需经过一个系统函数为  $H(z)=1-0.97z^{-1}$  的预加重处理来提升高频分量<sup>[10]</sup>。将语音信号进行分帧处理,这时每一帧可视为稳态信号,同时再对每帧信号乘以汉明窗来增加相邻帧的连续性,从而减少频域中的泄露。然后对每一帧信号做快速傅里叶变换,再计算其能量。把得出的每帧信号能量通过梅尔滤波器并计算每个滤波器组输出的对数能量,最后经过离散余弦变换 (Discrete Cosine Transform, DCT) 得到 MFCC 系数。

### 2.2 隐马尔科夫 (HMM)

隐马尔科夫模型 (HMM) 广泛地应用在语音处理等领域。一个 HMM 可由  $\lambda=\{M, O, A, B, \pi, F\}$  来描述。其中  $M$  为有  $N$  个状态组成的有限集合,  $O$  则是观察序列。  $A$ 、  $B$  分别为转移状态概率和输出观测值概率矩阵,  $\pi$  为初始概率序列,  $F$  则是终止状

态序列。HMM 语音识别首先根据前后向算法和递推算法对给定模型的输出和初始模型的输出序列的概率进行计算,再利用 Baum Welch 算法和最大似然准则对模型校正,最后利用维特比算法解码得到识别结果。本文采用连续混合密度 HMM, 由于一个高斯概率密度函数不能完整表征信号特征参数的概率分布,而任何一个函数都能通过多个高斯概率密度函数加权求和表示,所以用多个高斯概率分布加权求和来表示输出概率密度,对于非特定人的语音识别效果明显。选用自左向右无跨越的结构,此结构通常能更好地体现人的发音状态,能达到较好的识别效果。HMM 结构如图 3 所示。

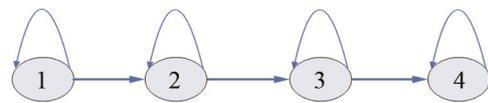


图 3 HMM 结构  
Fig.3 The structure of HMM

## 3 仿真实验以及结果分析

本次实验结合公共安防课题,解决传统视频监控只能看不能听的问题,利用语音识别技术对部分敏感词汇进行识别。数据使用 M-AUFIO 音频采集器完成,录制环境为相对安静的楼顶天台。噪声和语音分别在同样的环境下采集。本实验由 20 位同学参与录制,有 13 位男生和 7 位女生。其中随机抽取 10 人的语音 (400 条) 作为训练样本,另外 10 人的语音 (240 条) 做测试用。每人分别录制 12 个敏感词汇: 火灾、爆炸、抢劫、盗窃、中毒、溺水、晕倒、危险、救命、受伤、救护车、消防车。噪声采集使用三种,分别为白噪声、F16 和 volvo 噪声,本文语音和噪声的实验设备采样频率均为 44.1 kHz,采样精度为 32 bit。在实际仿真中经过了降采样处理,采样率为 16 kHz,帧长为 512,帧移是 256,窗函数为 Hamming 窗。特征参数采用 12 维的梅尔倒谱系数,选用连续混合密度 HMM,模型结构如图 3 所示,它包含 4 个状态,每状态含有 3 个高斯概率密度函数。为了验证本文方法的可行性,选取以下三种方法做对比。分别为: 谱减法、改进 Wiener 滤波、LogMMSE-IMCRA。图 4 为在 F16 噪声干扰下各算法增强后时域仿真波形,选用的语音内容为“救护车”,信噪比为 0 dB。

从图 4 可以看出,谱减法增强效果并不理想。LogMMSE-IMCRA 较改进 Wiener 有更明显的增强效果,但是两者在时域波形上表现出较多的毛刺和

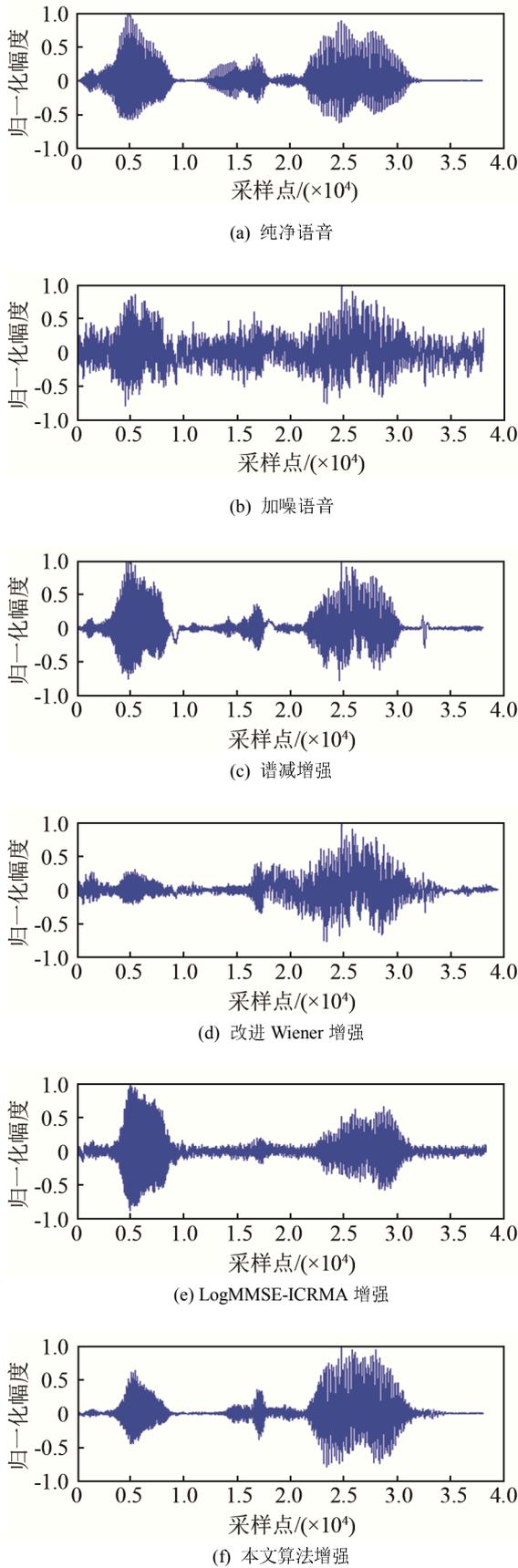


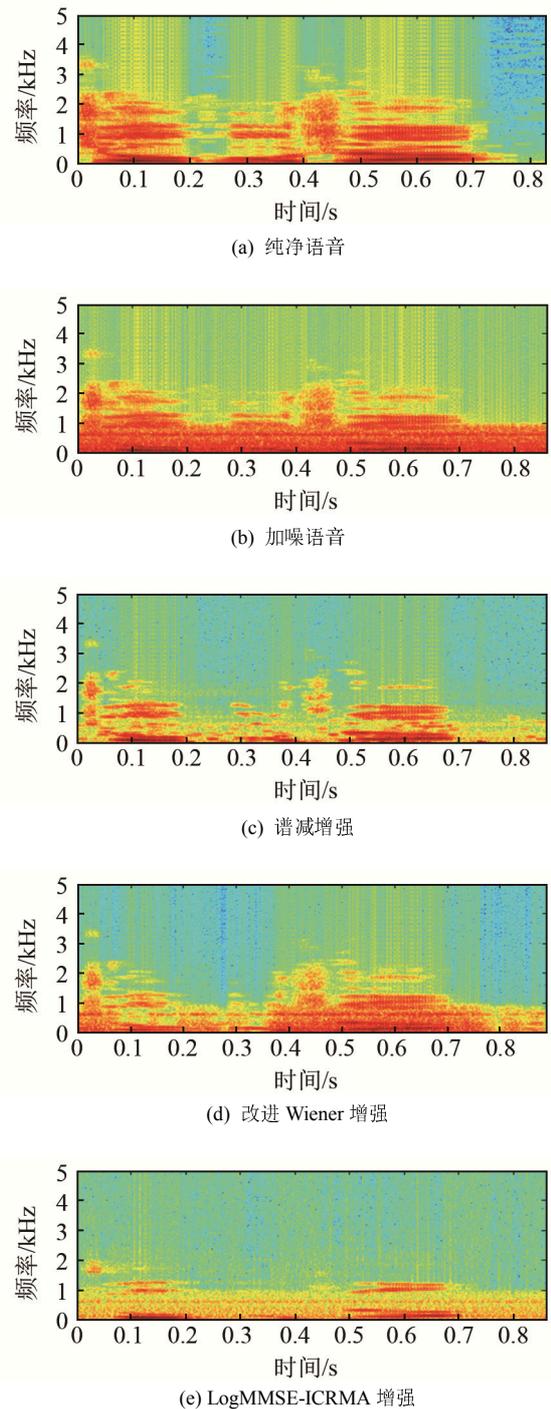
图4 F16 噪声环境下增强后的语音时域仿真图

Fig.4 The simulation diagrams of the speeches enhanced by different algorithms in F16 noise environment

噪声残留。本文所使用的增强方法效果明显，虽然仍会产生部分失真，但在低信噪比的环境下是可以接受的。

图 5 为增强后的语谱图，图 5 中所列的语谱图分别对应图 4 中的各时域仿真图。从语谱图来看，本文算法能更好地去除噪声，减少语音畸变，信号能量在低频段明显，增强后与原始语音基本保持一致。

图 6(a)、6(b)、6(c)分别是在白噪声、F16 和 volvo 噪声环境下通过四种算法增强后的识别率。



(e) LogMMSE-ICRMA 增强

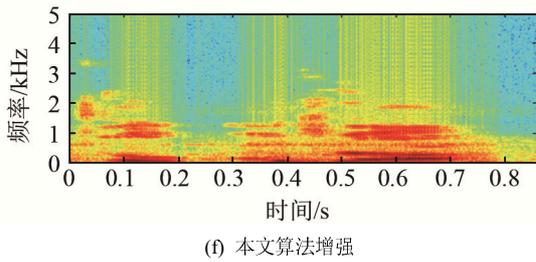


图 5 F16 噪声环境下增强后语谱图

Fig.5 The spectrograms of the speeches enhanced by different algorithms in F16 noise environment

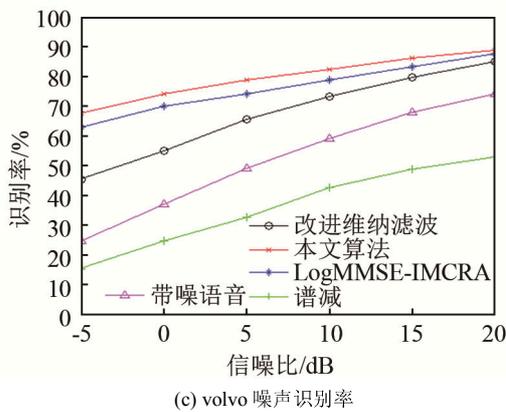
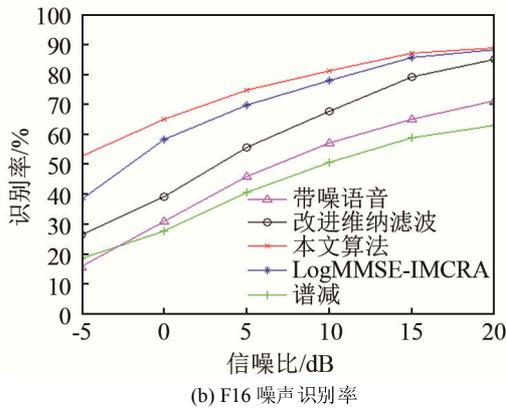
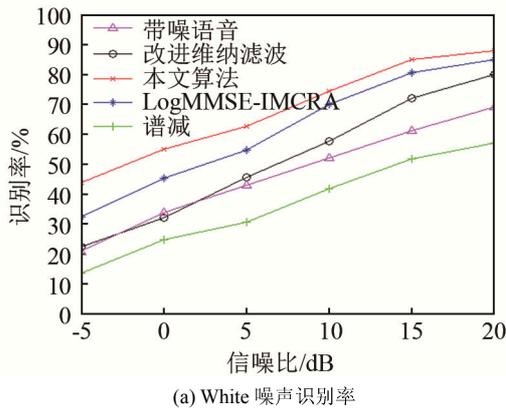


图 6 不同噪声环境下增强后的语音识别正确率

Fig.6 The accuracy rates of enhanced speech recognition in different noisy environments

从图 6 中对比可以看到，谱减法表现出较低的识别率，这是由于谱减法在增强后引入音乐噪声，使语音产生畸变从而降低识别率。改进 Wiener 滤波比谱减法的识别率要高，是因为其有更好的噪声抑制效果。LogMMSE-IMCRA 算法通过准确的噪声估计来增强语音能有效提高识别率，但由于去噪后仍有较多噪声残留，识别率表现一般。使用本文算法增强后的识别率明显提高，尤其在低信噪比环境下效果更为明显。这是因为本算法不仅对信号的增强去噪效果明显，还针对语音信号消噪后容易产生畸变失真的区域，对其进行增益补偿，最大限度地减少语音失真以提高识别率。

表 1 列举了三种噪声环境在 -5dB 环境下通过四种算法增强后的平均运行时间，单位是秒(s)。四种算法的运行环境均为在同一设备下运行，实验仿真平台为 matlab2012(a)，所使用的计算机 CPU 主频为  $2.40 \times 10^6$  kHz。从表 1 中可以看出，谱减法和改进 Wiener 虽然运行时间较短，但去噪效果并不明显。LogMMSE-IMCRA 由于引入了递归平均算法做噪声估计，所以运行时间较长，而本文算法与其他算法相比在减少了运算量的同时也达到了较好的去噪效果。

表 1 不同方法在 -5 dB 环境下语音增强的时间对比(s)

Table 1 The running time of enhanced speech in -5 dB SNR for different methods (s)

方法	White 噪声	F16 噪声	volvo 噪声
谱减法	0.185 8	0.165 2	0.181 4
改进 Wiener	0.756 1	0.775 4	0.701 4
LogMMSE-IMCRA	1.583 5	1.438 7	1.612 5
本文算法	1.125 9	1.072 6	1.218 6

表 2 列举了白噪声、F16 和 volvo 噪声在 -5 dB 环境下通过四种算法增强后的识别率。可以看出本文算法在 volvo 噪声下相比较其他两种噪声表现出更高的识别效果，这是由于基于 LogMMSE 算法在提升语音可懂度方面相对于传统增强算法在非平稳噪声中有更好的表现。而改进的 Wiener 滤波又能有效降低语音畸变，进一步提升语音可懂度。

表 2 不同方法在信噪比为 -5 dB 环境下语音增强后识别正确率对比(%)  
Table 2 The accuracy rates of enhanced speech recognition in -5 dB SNR environment for different methods (%)

方法	White 噪声	F16 噪声	volvo 噪声
谱减法	13.2	18.5	15.5
改进 Wiener	18.3	25.4	44.4
LogMMSE-IMCRA	32.7	38.3	63.3
本文算法	43.7	52.7	67.7

本文通过随机抽取不同噪声环境在 $-5$  dB 到  $5$  dB 的识别情况发现, 同一信噪比环境下的不同组, 所识别不正确的单词是不同的。也就是说低信噪比下识别错误的单词是随机的。通过提取识别错误的单词发现, 其时域波形表现为仍有部分噪声残留或失真, 所以导致识别错误, 这种情况随着信噪比的提高而减小。另一种原因可能是有的参与录制人员说话夹杂方言, 在理想环境下可以有效识别, 但经过处理后识别效果下降。

## 4 结 论

针对低信噪比环境下识别率不高的问题。本文先用 LogMMSE 提高各通道信噪比, 再利用改进的 Wiener 滤波去除噪声残留并降低由于增强处理所导致的语音畸变, 最后对增强后的信号进行识别。实验表明本文方法相比较 LogMMSE-IMCRA 算法不仅取得了更好的识别效果, 同时还减少了算法的计算量, 而且更适用于低信噪比环境中。

### 参 考 文 献

- [1] Loizou P C. Speech enhancement: theory and practice[M]. The Chemical Rubber Company Press, 2013: 75-109.
- [2] 胡丹, 曾庆宁, 龙超, 等. 连续语音识别前端鲁棒性研究[J]. 电视技术, 2015, 39(24): 43-46.  
HU Dan, ZENG Qingning, LONG Chao, et al. Front-end robust study for continuous speech recognition[J]. Video Engineering, 2015, 39(24): 43-46.
- [3] 曹亮, 张天骐, 高洪兴, 等. 基于听觉掩蔽效应的多频带谱减语音增强方法[J]. 计算机工程与设计, 2013, 34(1): 235-240.  
CAO Liang, ZHANG Tianqi, GAO Hongxing, et al. Multi-band spectral subtraction method for speech enhancement based on masking property of human auditory system[J]. Computer Engineering and Design, 2013, 34(1): 235-240.
- [4] Jose A Gonzalez, Antonio M Peinado, Ma N, et al. MMSE-Based missing-feature reconstruction with temporal modeling for robust speech recognition[J]. Audio Speech & Language Processing IEEE Transactions on, 2013, 21(3): 624-635.
- [5] Cohen I, Berdugo B. Speech enhancement for non-stationary noise environments[J]. Signal Processing, 2009, 81(11): 2403-2418.
- [6] 张东方, 蒋建中, 张连海. 一种改进型 IMCRA 非平稳噪声估计算法[J]. 计算机工程, 2012, 38(13): 270-272.  
ZHANG Dongfang, JIANG Jianzhong, ZHANG Lianhai. Improved IMCRA non-stationary noise estimation algorithm[J]. Computer Engineering, 2012, 38(13): 270-272.
- [7] 张亮, 龚卫国. 一种改进的(Wiener)滤波语音增强算法[J]. 计算机工程与应用, 2010, 46(26): 129-131.  
ZHANG Liang, GONG Weiguo. Improve wiener filtering speech enhancement algorithm[J]. Computer Engineering and Applications, 2010, 46(26): 129-131.
- [8] Fei C, Loizou P C. Impact of SNR and gain-function over- and under-estimation on speech intelligibility[J]. Speech Communication, 2012, 54(2): 272-281.
- [9] 郭利华, 马建芬. 具有高可懂度的改进的(Wiener)滤波的语音增强算法[J]. 计算机应用与软件, 2014(11): 155-157.  
GUO Lihua, MA Jianfen. An improved wiener filtering speech enhancement algorithm with high intelligibility[J]. Computer Applications and Software, 2014(11): 155-157.
- [10] 宋知用. MATLAB 在语音信号分析与合成中的应用[M]. 北京: 北京航空航天大学出版社, 2013.  
SONG Zhiyong. The application of MATLAB in speech signal analysis and synthesis[M]. Beijing: Beihang University Press, 2013.